Fusion Framework for Moving-Object Classification

R. Omar Chavez-Garcia*, Trung-Dung Vu*, Olivier Aycard* and Fabio Tango†
*University of Grenoble1, Grenoble - France
Email: {ricardo.chavez-garcia, trung-dung.vu, olivier.aycard}@imag.fr
†Centro Ricerche Fiat, Orbassano - Italy
Email: fabio.tango@crf.it

***Abstract*—Perceiving the environment is a fundamental task for Advance Driver Assistant Systems. While simultaneous localization and mapping represents the static part of the environment, detection and tracking of moving objects aims at identifying the dynamic part. Knowing the class of the moving objects surrounding the vehicle is a very useful information to correctly reason, decide and act according to each class of object, e.g. car, truck, pedestrian, bike, etc. Active and passive sensors provide useful information to classify certain kind of objects, but perform poorly for others. In this paper we present a generic fusion framework based on Dempster-Shafer theory to represent and combine evidence from several sources. We apply the proposed method to the problem of moving object classification. The method combines information from several lists of moving objects provided by different sensor-based object detectors. The fusion approach includes uncertainty from the reliability of the sensors and their precision to classify specific types of objects. The proposed approach takes into account the instantaneous information at current time and combines it with fused information from previous times. Several experiments were conducted in highway and urban scenarios using a vehicle demonstrator from the interactIVe European project. The obtained results show improvements in the combined classification compared with individual class hypothesis from the individual detector modules.**

I. Introduction

Machine perception is the process of understanding and representing the environment by organizing and interpreting information from sensors. Intelligent vehicles applications like the Advance Driver Assistant Systems (ADAS) help drivers to perform complex driving tasks and avoid dangerous situations. ADAS generally have three components: perception, reasoning & decision and control. We have to perceive from the sensors in order to model the current static and dynamic environment. Then, we use the perception output to reason and decide which actions are the best to finally perform such actions. In order to perform a good reasoning and control we have to correctly model the surrounding environment [1].

Robotic perception is composed of two main tasks: simultaneous localization and mapping (SLAM) deals with modeling static parts; and detection and tracking moving objects (DATMO) is responsible for modeling dynamic parts of the environment. In SLAM, when vehicle location and map are unknown the vehicle generates a map of the environment while simultaneously localizing itself in the map given all the measurements from its sensors. DATMO aims at detecting and track the moving objects surrounding the vehicle and predict their future behaviors. SLAM and DATMO are considered correlated and aim at obtaining a holistic representation of the environment: static and moving objects [1], [2].

Once object detection and tracking is done a classification step is needed in order to determine which class of objects are surrounding the vehicle. Knowledge about the class of surrounding moving objects can help to improve their tracking, reason about their behavior and decide what to do according to their nature. Moving object classification enables a better understanding of driving situations. Object detectors that aim to recognize specific type of objects are considered binary classifiers.

Current state of the art approaches for object classification focus only in one class of object (e.g. pedestrians, cars, trucks, etc.) and rely on one type of sensor (active or passive) to perform such task. Including information from different type of sensors can improve the object classification and allow the classification of multiple class of objects [3]. Individual object classification from specific sensors, like camera, lidar or radar, have different reliability degrees according to the sensor advantages and drawbacks. Therefore, we use the ability of each sensor to compensate the deficiencies of others and hence to improve the final classification.

Fusion approaches for object classification should aim to complement the sensor advantages and reduce their disadvantages to improve the classification result [4]. It is discussed in [1], [5] that, for classification purposes, when combining different sensor inputs we must be aware about the classification precision of each sensor and take it into account to get a more accurate result.

Geronimo et al. review the current state of the art in pedestrian detection for ADAS, their work focus on camera-sensor based approaches due to the high potential of visual features, spatial resolution and richness of texture and color cues [3]. They conclude that the advantages and drawbacks of camera sensors can be complemented with active sensors, like lidar, to improve the overall performance of pedestrian detection.

Vehicle detection based only on camera sensors is very challenging due to big intraclass differences like shape, size and color. Changes in illumination, object's pose, and surrounded objects make difficult the appearance identification. The works proposed in [6] and [7] review different approaches for vehicle detection, it highlights the difficult of relying only on camera sensors and suggest that the use of active sensors could improve the detection performance.

Himmelsbach [8] and Azim [9] propose a method for tracking and classifying arbitrary objects using active sensors, their proposed methods show good results. Himmelsbach proposes a bottom-up/top-down approach that relies on a 3D lidar sensor and considers object appearance and motion history in the classification task. Its main limitation is that it needs top-down knowledge to properly perform the cloud of points classification. Azim proposes a three dimensional representation of the environment to perform the moving object detection and classification, but its results show a limited performance in cluttered environments due to the lack of discriminative information.

Smets in [10] proposes an approach to joint tracking and classification using Dempster-Shafer theory. This approach uses classical Kalman Filters for the tracking phase while for the classification part it uses the transferable belief model (TBM) framework. Results show that when there is no one-to-one mapping between target behaviours and classes, TBM provides more intuitive results than a Bayesian classifier. However, this approach does not take into account multiple sources of evidence which can help to improve the tracking and classification accuracy; and moreover, relies entirely on the targets behaviours to perform the classification without taking into account appearance information.

While in probability theory, evidence of a variable value may be placed on any element of a possible set of values. In Dempster-Shafer (DS) theory, evidence cannot only be placed on elements and sets, but also on sets of sets. This means that the domain of DS theory is all sets of all subsets. DS theory provides tools for capturing ignorance or an inability to distinguish between alternatives. In probability theory, this would be done by assigning an equal or uniform probability to each alternative [11]. The use of the power set as the frame of discernment allows a richer representation of hypothesis. Using combination rules we can fuse the evidence from different sources to transfer the evidence into a final combined result [12].

In this paper we propose a generic fusion approach based on DS theory. We apply this approach to the moving object classification problem. Given a list of detected objects and a preliminary classification from different individual detectors (or classifiers), the proposed approach combines instantaneous information from current environment state by applying a rule of combination based on the one proposed in [13]. The rule of combination can take into account classification evidence from different sources of evidence (object detectors/classifiers), the uncertainty coming from the reliability of the sensors and the sensor's precision to detect certain classes of objects. The proposed approach aims to improve the individual object classification provided by class-specific sensor detectors. After instantaneous fusion is done the proposed approach fuses it with the combination result from previous times. Its architecture allows to give more importance to the classification evidence according to its uncertainty factors.

Using the DS theory we are able to represent the evidence coming from different sensors into a common representation based on prepositions, i.e. object class hypothesis. The proposed fusion framework relies in two main parts: the instantaneous fusion obtained from sensor evidence at current time, and the combined evidence from previous times. Instantaneous fusion is divided in two main layers: the individual evidence layer and the fusion layer. This features allow the proposed method to include several sensor inputs and different sets of object classes.

The main contributions of this work rely on: the definition of a generic fusion framework based on DS theory and particularly applying to the moving object classification task; the inclusion of sensors reliability and sensors precision to detect and classify certain classes of objects as main parameters to perform fusion; a conjunctive rule of combination that allows to combine several sources of evidence using a common representation frame, it allows as well to include uncertainty from the reliability and sensor classification precision, and to manage conflict situations that can lead counter-intuitive results.

Several experiments were done to analyze and compare the results obtained by the proposed fusion approach against the individual classification provided by three individual object detectors. We used real data from highways and urban areas obtained by a demonstrator from the interactIVe (Accident Avoidance by Active Intervention for Intelligent Vehicles) European project [1].

This rest of this paper is organized as follows. Next section reviews some concepts of the Dempster-Shafer theory. Section III describes the vehicle demonstrator and the set of sensors we use to test our proposed approach. In section IV we define the proposed fusion framework used to combine classification information. Implementation of the proposed fusion framework is done using the architecture define in section V. Experimental set-up and experimental results are shown in section VI. Finally, section VII presents the conclusions.

II. Dempster-Shafer theory background

The Dempster-Shafer theory is a generalization of the Bayesian theory of subjective probability. Whereas the Bayesian theory requires probabilities for each question of interest, DS theory allows us to base degrees of belief for one question on probabilities for a related question [12]. DS theory is highly expressive, allows to represent different levels of ignorance, does not need prior probabilities and manage conflict situations when opposite evidence appears.

DS theory represents the world in a set of mutually exclusive propositions known as the frame of discernment (Ω). It uses belief functions to distribute the evidence about the propositions over 2^{Ω}. The distribution of mass beliefs is done by the function $m : 2^{\Omega} \rightarrow [0, 1]$, also known as Basic Belief Assignment (BBA), which is described in equation 1.

DS theory allows alternative scenarios other than the single hypotheses, such as considering equally the possible sets that have a non-zero intersection. We can combine hypothesis

[1] http://www.interactive-ip.eu

in a compound set giving it a new semantic meaning, for example the *unknown* hypothesis created from combining all the individual hypothesis. Moreover BBA can supports any proposition $A \subseteq \Omega$ without supporting any sub-proposition of A, which allows to represent partial knowledge.

$$\begin{aligned} &m(\emptyset) = 0; \\ &\textstyle\sum_{A \subseteq \Omega} m(A) = 1. \end{aligned} \tag{1}$$

Any subset A of Ω with $m(A) > 0$ for a particular belief function is called a focal element of that function.

In order to combine different sources of evidence, represented as belief functions over the same frame of discernment and with at least one focal element in common, a combination rule is required. Several fusion operators have been proposed into the DS framework concerning scenarios with different requirements. One of the widely used is that proposed by Dempster [12]. Dempster's rule of combination assumes independence and reliability of both sources of evidence. This rule is defined as follows:

$$\begin{aligned} m_{12}(A) &= \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - K_{12}}; \; A \neq \emptyset \\ K_{12} &= \textstyle\sum_{B \cap C = \emptyset} m_1(B) m_2(C) \end{aligned} \tag{2}$$

where K_{12} is known as the degree of conflict. Dempster's rule analyses each piece of evidence to find conflicts and uses it to normalize the masses in the set.

III. Vehicle Demonstrator

We use the CRF vehicle demonstrator, which is part of the European project interactIVe, to obtain datasets from highway and cluttered urban scenarios. The obtained datasets were used to test our proposed fusion framework. The demonstrator is a Lancia Delta car equipped from factory with electronic steering systems, two ultrasonic sensors located on the side of the front bumper, and with a front camera located between the glass and the central rear mirror. Moreover, the demonstrator vehicle has been installed with a scanning laser and a mid-range radar on the front bumper for the detection of obstacles ahead, as depicted in figure 1. Finally, two radar sensors have been installed on both sides of the rear bumper to cover the side and rear areas.

IV. Fusion Approach for Moving-Object Classification

This work proposes an information fusion framework which allows to incorporate in a generic way information from different sources of evidence. This fusion approach is based on DS theory and aims to gather classification information from moving objects identified by several detector modules. These detector modules can use information from different kind of sensors. This proposed approach provides as output a fused list of classified objects.

Figure 2 shows the general architecture of the proposed fusion approach. The input of this method is composed of several lists of detected objects and their class information,

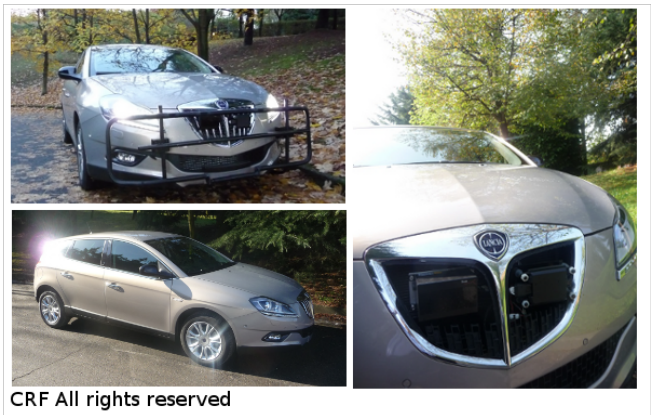

Fig. 1. Images of the CRF demonstrator.

the reliability of the sources of evidence and the precision detection for certain type of classes. We assign empirically an evidence value to each object (set element) regarding these last two factors. Using a proposed conjunctive rule of combination, we combine the classification information from detector modules at a current time to obtain an instantaneous combination, later on this instantaneous class information is fused with previous combinations in a process we call dynamic combination. The final output of the proposed method comprise a list of objects with combined class information.

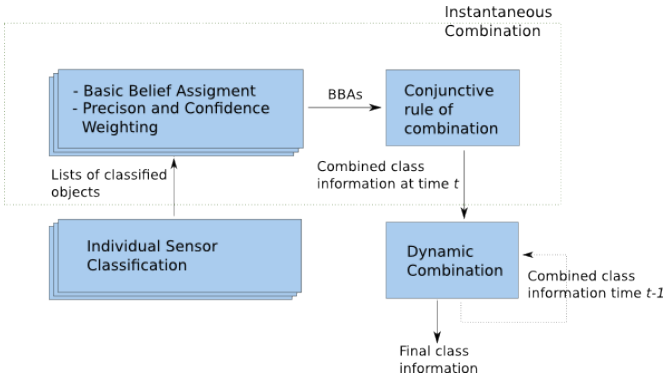

Fig. 2. Schematic of the proposed fusion architecture.

A. Instantaneous Combination

According to Dempster-Shafer's theory, let's define the frame of discernment Ω and the power set of Ω as the set of all possible class hypothesis for each source of evidence. Where Ω represents the set of all the classes we want to identify. Let's define m_b as the reference mass evidence and m_c as the mass evidence from the source we want to combine with. Finally, m_a represents the combined mass evidence.

In situations where the conflict mass is high, Dempster's combination rule generates counter-intuitive results, for this reason we decide to adapt the combination rule proposed by Yager in [13] to obtain a most suitable rule of combination that will avoid this counter-intuitive results moving the conflict

mass (K) to the set Ω. This means transferring the conflict mass to the ignorance state instead of normalizing the rest of the masses. We do this expecting that future mass evidence will help to solve conflict states when two unreliable sources of evidence classify differently one object. The used rule of combination is stated as follows.

$$\begin{aligned} m_r(A) &= \textstyle\sum_{B\cap C=A} m_b(B)m_c(C);\ A \neq \emptyset \\ K &= \textstyle\sum_{B\cap C=\emptyset} m_b(B)m_c(C) \\ m_r(\Omega) &= m'_r(\Omega) + K \end{aligned} \tag{3}$$

Where $m'_r(\Omega)$ is the BBA for the ignorance state and $m(\Omega)$ includes the added ignorance from the conflict states. This rule considers both sources of evidence are independent and reliable.

As we cannot assure the reliability of the evidence sources regarding the classification due to sensor limitations or miss classifications, we proposed to use a *discounting* factor for each source of evidence [14]. We believe doing this will allow us to overcome this issue.

Let's define m_a as a reliable reference source of evidence and m_b as a relative reliable source of evidence. We define $r_{ab} \in [0, 1]$ as the reliability factor of m_b with respect to m_a. To make m_b reliable we apply r_ab over the BBA of m_2. The evidence we take from the subsets of 2^Ω after applying the reliability factor should be consider ignorance, therefore is transfer to the set Ω

$$\begin{aligned} m_b(A) = &\quad r_{ab} \times m'_b(A); A \subseteq 2^\Omega, A \neq \Omega \\ m_b(\Omega) = &\quad m'_b(\Omega) + \textstyle\sum(1 - r_{ab} \times m(A)); \\ &\quad for A \subseteq 2^\Omega, A \neq \emptyset, A \neq \Omega \end{aligned} \tag{4}$$

This means that we adjust the mass evidence of m_b according to how reliable it is compared with the reference source of evidence m_a. When m_b is as reliable as m_a ($r_{ab} = 1$) we get the original BBA for m'_b:

$$\begin{aligned} m_b(A) &= m'_b(A) \\ m_b(\Omega) &= m'_b(\Omega) \end{aligned} \tag{5}$$

There are scenarios where one of the sources of evidence is more precise to identify the class of an specific subset of the frame of discernment. We can include this uncertainty using a similar approach to the one prosed above for the reliability but focus in specific subsets of the frame of discernment.

Let's $f_i \in [0, 1]$ be the precision factor for the ith subset (hypothesis) of a particular belief function m_a. The greater the value the more precise is the source evidence about the mass evidence assign to the subset.

$$\begin{aligned} m_a(A_i) = &\quad m'_a(A_i) \times f_i; A_i \subseteq 2^\Omega, A_i \neq \emptyset \\ m_a(\Omega) = &\quad m'_\Omega + \textstyle\sum(1 - f_i) \times m'_a(A_i); \\ &\quad for A_i \subseteq 2^\Omega, A_i \neq \emptyset, A_i \neq \Omega \end{aligned} \tag{6}$$

Where m'_a represents the reliable BBA. All the unallocated evidence will be placed in the Ω state because it is considered ignorance.

Once we have applied the reliability and precision factors, the combination rule in equation 3 can be used. Several sources can be combined applying iteratively this rule of combination and using the fused evidence as the reliability reference source.

The final fused evidence contains the transferred evidence from the different sources. The criterion we use to determine the final hypothesis is based on the higher mass function value from the combined set, though it can be modified to be based on $belief$ or $plausibility$ degrees.

B. Dynamic Combination

Since we are performing the combination of different sources of evidence at a current time t, we will call this instantaneous fusion. One can notice that including information from previous combination can add valuable evidence to the current available evidence. Regarding this topic, and taking advantage of the proposed general framework architecture, we introduce equation 7 as an extension of the proposed instantaneous fusion to include mass evidence from previous combinations (e.g. time $t - 1$).

$$m_{r^t}(A) = m_r(A) \otimes m_{r^{t-1}}(A) \tag{7}$$

Where $m_r(A)$ represents the instantaneous fusion at time t. The operator $\otimes$ follows the same combination rule defined in equation 3, which is used as well to obtain the instantaneous fusion. Following this extension we can notice that the combined mass for the list objects from all the previous times is represented in $m_{r^{t-1}}(A)$.

V. Frontal Object Perception

Figure 3 shows the general frontal object perception (FOP) architecture used in the interactIVe project for the vehicle demonstrator described in section III. FOP takes raw information from three different sensors to detect static and moving objects in the surrounding environment. While lidar processing detects and track moving objects, pedestrian and vehicle detectors focus on regions of interest provided by lidar processing to provide more classification evidence. The three detector modules provided a list of moving objects and their preliminary classification. The objective of the proposed fusion approach defined in section IV is to take the three classification hypothesis provided by the three object detectors, to combine them and obtain a final classification for each moving object. The proposed fusion approach focus only in the class information provided by the object detector modules. Raw data processing for objects detection and moving object tracking is performed by the frontal object perception module described in detail in [15].

Lidar Target Detection and Tracking

Raw lidar scans and vehicle state information are processed to recognized static and moving objects, which will be maintained for tracking purposes. We employ a grid-based fusion

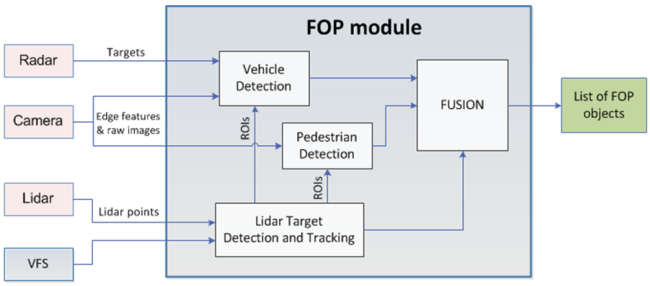

Fig. 3. General architecture of the FOP module for the CRF demonstrator.

approach originally presented in [16] and which incrementally integrates discrete lidar scans into a local occupancy grid map representing the environment surrounding the ego-vehicle. In this representation, the environment is discretized into a two-dimensional lattice of rectangular cells; each cell is associated with a measure indicating the probability that the cell is occupied by an obstacle or not. A high value of occupancy grid indicates the cell is occupied and a low value means the cell is free. By using this grid-based representation noise and sparseness of raw lidar data can be inherently handled, moreover no data association is required. We analyze each new lidar measurement to determine if static or moving objects are present.

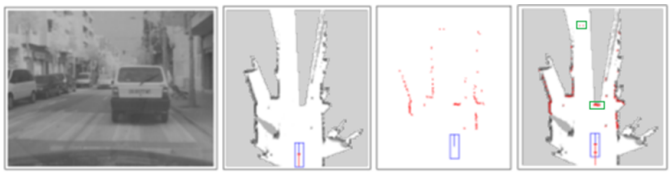

Fig. 4. Occupancy grid representation obtained by processing raw lidar data. From left to right: Reference image from camera; occupancy grid obtained by combining all raw lidar measurements; local occupancy grid from latest lidar measurements; and identification of static and dynamic objects (green bounding boxes).

By projecting the impact points from latest laser measurements onto the local grid, each impact points is classified as static or dynamic. The points observed in free space are classified as dynamic whereas the rest are classified as static. Using a clustering process we identify clouds of points that could describe moving objects. Then the list of possible moving objects is passed to the tracking module. Figure 4 shows an example of the evolution of the moving object detection process using an occupancy grid representation.

We represent the moving objects of interest by simple geometric models, i.e.: rectangle for vehicles and bicycles, small circle for pedestrians. Considering simultaneously the detection and tracking of moving objects as a batch optimization problem over a sliding window of a fix number of data frames, we follow the work proposed by Vu et al. [17]. It interprets the laser measurement sequence by all the possible hypotheses of moving object trajectories over a sliding window of time. Generated object hypotheses are then put into a top-down process (a global view) taking into account all object dynamics model, sensor model and visibility constraints. The data-driven Markov chain Monte Carlo (DDMCMC) technique is used to sample the solution space effectively to find the optimal solution.

Vehicle and Pedestrian Detectors

Regions of interest, provided by lidar processing, are taken by the vehicle and pedestrian detectors to perform the camera-based object classification. Vehicle detector uses both radar and video outputs at two different steps to perform robust vehicle detection. Radar sensors have a good range resolution and a crude azimuth estimation and camera sensors are able to give a precise lateral estimation while having an uncertain range estimation. The pedestrian detector module detailed in [15] scans the image using a sliding window of fixed size to detect the pedestrians. For each window, visual features are extracted and a classifier (trained off-line) is applied to decide if an object of interest is contained inside the window. A modified version of histogram of oriented gradients (called sparse-HOG) features, which focus on important areas of the samples, powers the pedestrian and vehicle representations at training and detection time. Given computed features for positive and negative samples, we use the discrete Adaboost approach proposed in [18] to build the vehicle and pedestrian classifiers. Its trade-off between performance and quality makes it suitable for real-time requirements. The idea of boosting-based classifiers is to combine many weak classifiers to form a powerful one where weak classifiers are only required to perform better than chance hence they can be very simple and fast to compute.

Several images, extracted from the Daimler Pedestrian Benchmark data sets and from manually label data sets, were used as training samples to build the vehicle and pedestrian classifiers. Despite the drawbacks of a vision-based detector, we experimentally notice that it can better identify pedestrians.

Figure 5 shows an output example from pedestrian and vehicle detectors. Pedestrian and Vehicle detector outputs consist on a list of classified objects or region of interest. These regions were provided by the lidar processing as inputs and after performing specific class detections they determined if the region was a vehicle (car, truck), a pedestrian, a bike or an unknown object.

Classification Information Fusion

The list of objects provided by the lidar processing, vehicle and pedestrian detector are taken as inputs for the proposed classification fusion framework. The fusion process take into account the reliability of the sensors and their precision to identify a particular kind of objects. This fusion is done in two steps: first it performs and instantaneous fusion using the individual classification from the different sensors at the current time; secondly, it performs a dynamic fusion between the classification information obtained from previous frames and the instantaneous fusion result. The result of the final fusion is store over the time. This process will output a final list of objects plus a likelihood of the class of each object.

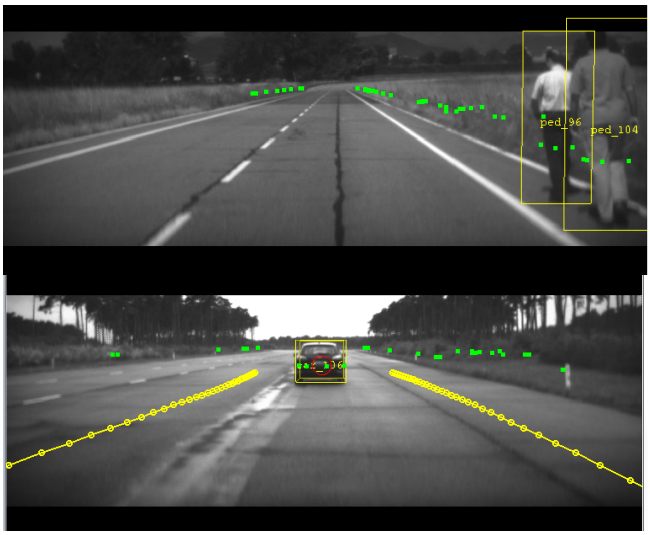

Fig. 5. Output example from pedestrian (top) and vehicle detector (down) after processing the inputs: regions of interest.

VI. Experiments And Results

Experiments were conducted using four short datasets provided by the vehicle demonstrator described in section III. We tested our approach in two main scenarios: urban and highway. The objective of these experiments was to verify if the results from our proposed approach improves the preliminary classification results provided by the individual object detector modules.

First of all we need to define the frame of discernment $\Omega = \{car, truck, pedestrian, bike\}$ and therefore the set of all possible 2^{Ω} classification hypothesis for each source of evidence, i.e. object detector output.

Following the fusion scheme presented in figure 2 and the general architecture from figure 3 we processed the individual list of preliminary classified objects provided by three different detector modules to obtain three lists of BBAs. Each individual list of objects contains either static or moving objects. The proposed fusion approach will focus on the moving objects. BBAs where defined empirically after analyzing the results from individual object detectors on datasets with none, one or several objects of interest. The reliability and precision factors where chosen according to the performance of each sensor processing on datasets from real driving scenarios.

Lidar target detector is able to identify all classes of objects using the cloud of points and the model-based approach. It has a good performance for identifying cars and trucks but a poor performance when it comes to pedestrians or bikes. We represent this behavior by setting individual precision factors as is shown in equation 6. While the precision factor is high for cars and trucks it is low for pedestrians and bikes. The uncertainty of the object class, for lidar detector, decreases when more information (frames) are available. This means, for example, that when the current lidar data indicates that a moving object is car this can be either a car or a truck, and vice-versa. When a car or truck is detected we set a high mass value in the respective individual set $\{car\} or \{truck\}$ and we split the remaining evidence into the $\{car, truck\}$ and Ω set. If a pedestrian or bike is detected, we perform the same mass assignment process, but according to the individual precision factors it will decrease the mass value in the individual sets $\{pedestrian\}$ and $\{bike\}$.

For each region of interest provided by lidar target detector, vehicle detector identifies a car, a truck or none of them. One can notice that some parts of vehicles are very similar, for example the rear part of the car or truck. For this reason there is uncertainty in the vehicle detection result. When a car is detected we put most of the evidence in the hypothesis car, the rest of the evidence is placed in the hypothesis that says the object could be a car or $truck$ and the ignorance set Ω. We use the same evidence assignation when a truck is detected. If no vehicle is detected in the region of interest, we put all the evidence in the ignorance set Ω.

Pedestrian detector's belief assignment is done in a similar way as with vehicle detector. When a pedestrian is detected we put high mass value in the hypothesis $pedestrian$ and the rest of evidence divided into the set $\{pedestrian, bike\}$ and the ignorance set Ω.

Each time we perform a combination between two BBA we have to define reliability factors relative to the current reference evidence source. Firstly, we set the lidar BBA as the reference evidence source and combined with the vehicle detector BBAs. Secondly, we set the combined BBA as the reference evidence before combine it with the pedestrian detector BBA to obtain the instantaneous fusion. For the final combination step, we set the combined mass from previous times as the reference evidence source and fuse it with the current instantaneous fusion. Individual precision factors are defined only for the individual BBAs.

Figure 6 shows the results of the fusion approach for moving object classification. We tested our proposed approach in several urban and highway scenarios. We obtained good results in both scenarios compared with the individual classification inputs. We are currently conducting several test in order to have quantitative values of the improvements achieved.

Figure 6 (a) shows how the proposed approach identifies the class of the two moving objects present in the scene: a car and a bike. In the contrary, in figure 6 (b) one pedestrian is missing because none of the object detector modules provided evidence to support its class.

The car in figure 6 (b) and the truck in figure 6 (c) are not classified by the lidar processing because they have just appeared few frames before in the field of view, but using the evidence derived from the lack of classification and from the vehicle detector the proposed approach can correctly identify both at early time. Mass evidence supporting this two classification hypothesis becomes higher in posterior frames when lidar processing provides evidence about the correct class of objects.

Figure 6 (d) shows how despite of the lack of texture in the image to identify the two vehicles in the front, evidence from the lidar processing helps to correctly identify them.

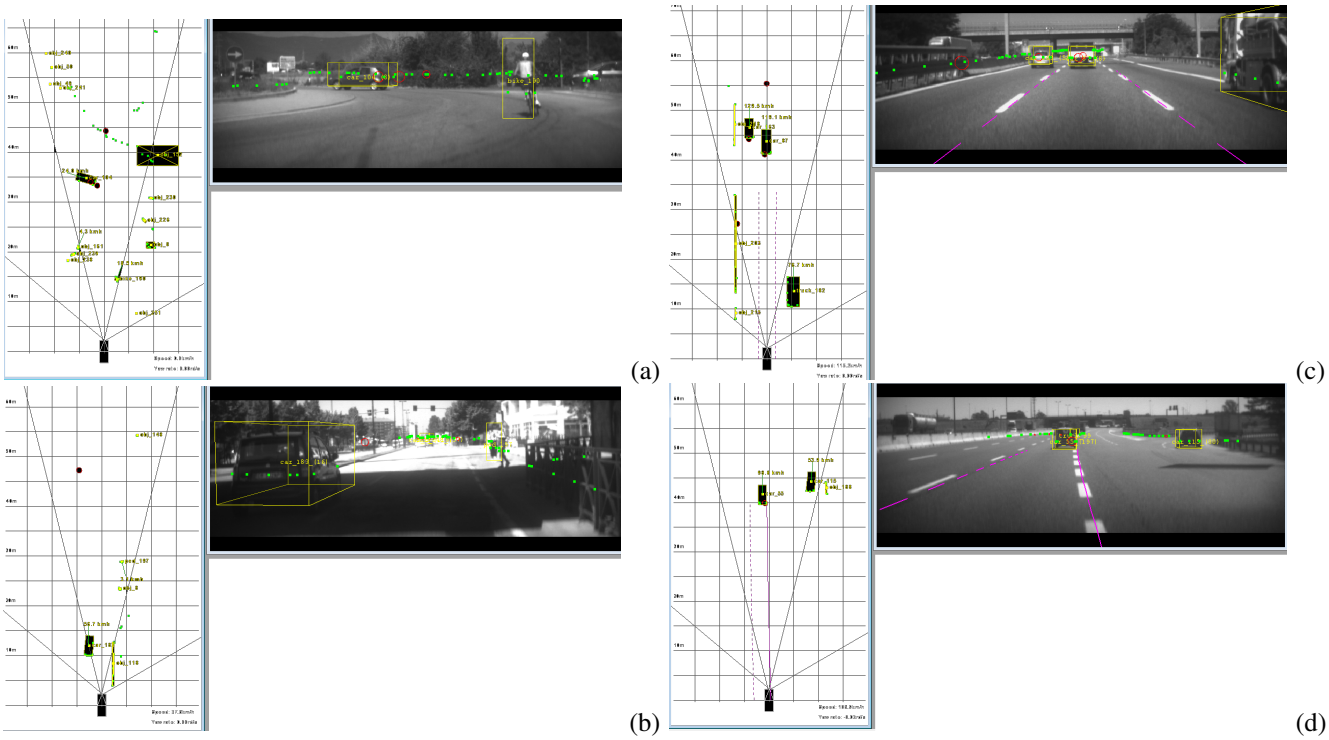

Fig. 6. Results from the proposed fusion approach for moving object classification in (a,b) urban and (c,d) highway scenarios. Left side of each image represents the top view representation of the scene (static and moving objects) showed in the right-side image. Bounding shapes and class tags in the right side of each image represent classified moving objects.

TABLE I
NUMBER OF VEHICLE/TRUCK MISS-CLASSIFICATIONS FROM INDIVIDUAL MOVING OBJECT DETECTORS AND FROM THE PROPOSED FUSION APPROACH.

Dataset	Lidar processing	Vehicle detector	Fusion approach
highway 1	9	10	4
highway 2	8	12	6
urban 1	15	19	10
urban 2	18	23	12

TABLE II
NUMBER OF PEDESTRIAN/BIKES MISS-CLASSIFICATIONS FROM INDIVIDUAL MOVING OBJECT DETECTORS AND FROM THE PROPOSED FUSION APPROACH. HIGHWAY DATASETS DO NOT CONTAIN ANY PEDESTRIAN OR BIKE.

Dataset	Lidar processing	Pedestrian detector	Fusion approach
urban 1	10	8	5
urban 2	13	7	3

Tables I and II show a comparison between the results obtained by the proposed fusion approach and the individual sources of evidence for moving object classification regarding the total number of object miss-classifications per dataset. We used four short datasets (about 1 minute long each one) provided by the vehicle demonstrator described in section III, two of them from highway scenarios and two from urban scenarios. We can see that the proposed fusion approach reduce the number of miss-classifications in all the datasets. Due to the specific nature of the pedestrian and vehicle detectors we divided the experiments to analyse separately the improvement in pedestrian/bike and car/truck classifications. One can see how different are the individual performances of the evidence sources to detect specific class objects, this is highly related to the nature of the sensor information they use to provide class hypotheses. The proposed fusion approach combines the class evidence provide by each source of evidence among its reliability and precision factors to obtain a better classification of the moving objects.

VII. CONCLUSIONS

In this work we presented a generic fusion framework based on DS theory to combine the class information of lists moving objects provided by different object detectors. The architecture of the proposed approach allows to include several sources of evidence as inputs despite their nature. Given the performance of the object detectors varies according to the type of sensor used and their specifications, we included two factors in the mass belief assignment: reliability of the sources and their precision to classify certain type of objects. Finally, we used a rule of combination to fuse several sources of evidence and manage situations with high levels of conflict without getting counter-intuitive results. These situations appear very often when different kind of sensors are used and when the ego-vehicle is placed in cluttered scenarios. Finally, the proposed fusion approach is able to fuse information from previous

combinations with the combined class information at current time (instantaneous fusion).

We used several datasets from urban and highway scenarios to test our proposed approach. These experiments showed improvements of the individual moving objects classification provided by lidar, pedestrian and vehicle object detectors. It is important to mention that these are preliminary results since the interactIVe project is still in its evaluation phase. We are currently developing an standard ground-truth dataset to evaluate the whole perception platform and therefore conduct experiments to obtain more quantitative results to support the improvements of the proposed classification fusion approach presented in this paper.

ACKNOWLEDGMENT

This work was also supported by the European Commission under interactIVe, a large scale integrating project part of the FP7-ICT for Safety and Energy Efficiency in Mobility. The authors would like to thank all partners within interactIVe for their cooperation and valuable contribution.

REFERENCES

[1] T.-D. VU, "Vehicle perception : Localization , mapping with detection , classification and tracking of moving objects," Ph.D. dissertation, LIG - INRIA, 2009.

[2] C. Wang, C. Thorpe, S. Thrun, M. Hebert, and H. Durrant-Whyte, "Simultaneous localization, mapping and moving object tracking," *The International Journal of Robotics Research*, vol. 26, no. 9, pp. 889–916, 2007.

[3] D. Gerónimo, A. M. López, A. D. Sappa, and T. Graf, "Survey of pedestrian detection for advanced driver assistance systems," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 7, pp. 1239–58, July 2010.

[4] M. Darms, P. Rybski, and C. Urmson, "Classification and tracking of dynamic objects with multiple sensors for autonomous driving in urban environments," *2008 IEEE Intelligent Vehicles Symposium*, pp. 1197–1202, June 2008.

[5] Q. Baig, "Multisensor data fusion for detection and tracking of moving objects from a dynamic autonomous vehicle," Ph.D. dissertation, University of Grenoble1, 2012.

[6] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection using optical sensors: a review," *Proceedings. The 7th International IEEE Conference on Intelligent Transportation Systems (IEEE Cat. No.04TH8749)*, pp. 585–590, 2004.

[7] M. Rohrbach, M. Enzweiler, and D. Gavrila, "High-level fusion of depth and intensity for pedestrian classification," in *DAGM-Symposium*, J. Denzler, G. Notni, and H. SuBe, Eds. Springer, 2009, pp. 101–110.

[8] M. Himmelsbach and H. Wuensche, "Tracking and classification of arbitrary objects with bottom-up / top-down detection," in *Intelligent Vehicles Symposium (IV)*, 2012, pp. 577– 582.

[9] A. Azim and O. Aycard, "Detection, classification and tracking of moving objects in a 3d environment," *2012 IEEE Intelligent Vehicles Symposium*, pp. 802–807, Jun. 2012.

[10] P. Smets and B. Ristic, "Kalman filter and joint tracking and classification based on belief functions in the tbm framework," *Information Fusion*, vol. 8, no. 1, pp. 16 – 27, 2007.

[11] Q. Baig, O. Aycard, T. D. Vu, and T. Fraichard, "Fusion between laser and stereo vision data for moving objects tracking in intersection like scenario," in *IEEE Intelligent Vehicles Symp.*, Baden-Baden, Allemagne, Jun. 2011.

[12] P. Smets, "The transferable belief model for belief representation," *In Gabbay and Smets*, vol. 6156, no. Drums Ii, pp. 1–24, 1999.

[13] R. R. Yager, "On the dempster-shafer framework and new combination rules," *Information Sciences*, vol. 41, no. 2, pp. 93 – 137, 1987.

[14] P. Smets, "Data fusion in the transferable belief model," in *Information Fusion, 2000. FUSION 2000. Proceedings of the Third International Conference on*, vol. 1, July 2000, pp. PS21–PS33 vol.1.

[15] O. Chavez-Garcia, J. Burlet, T.-D. Vu, and O. Aycard, "Frontal object perception using radar and mono-vision," in *2012 IEEE Intelligent Vehicles Symposium (IV)*. Alcalá de Henares, Espagne: IEEE Conference Publications, 2012, pp. 159–164, poster.

[16] A. Elfes, "Using occupancy grids for mobile robot perception and navigation," *Computer*, vol. 22, pp. 46–57, June 1989.

[17] T.-D. Vu and O. Aycard, "Laser-based detection and tracking moving objects using data-driven markov chain monte carlo," in *ICRA*, 2009, pp. 3800–3806.

[18] R. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," *Machine learning*, 1999.